

**Implementation of the  
Open Archive Initiative Protocol for Metadata Harvesting  
Version 2.0 in NASA Digital Library Repositories**

---

Terry L. Harrison  
Old Dominion University  
Langley Aerospace Research Summer Scholars (LARSS) Program  
NASA/ Langley Research Center, Hampton, Virginia

---

Mentor: Michael L. Nelson, Ph.D.  
Systems Engineering

08.06.2002

**Abstract:**

Digital libraries are critical to the dissemination of knowledge. They are used at NASA to provide a point of access, via the World Wide Web, to publicly available information, including its technical reports. Digital library technologies are implemented at NASA Langley Research Center (LaRC) in accordance with the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) to establish a collection of data and service providers.

In June of 2002, the OAI released the 2.0 version of the OAI-PMH, which made it necessary to recode NASA LaRC data providers in order for them to maintain their compliance with the protocol. During the 10 weeks of the LARSS program, Terry Harrison has implemented the new OAI-PMH for all of NASA's data providers. In addition, he has successfully deployed code to add data provider functionality to the NASA Technical Reports Server, permitting a service provider to now operate as an metadata aggregator for NASA's collection of publicly available technical reports.

While focusing on the distribution of NASA's public technical reports, the development of effective metadata harvesting has great implications for widespread application to all digital resources.

**Introduction:**

NASA was created by Congress in 1958 with the National Aeronautics and Space Act, which required by law, to "provide for the widest practical and appropriate dissemination of information concerning its activities and results". In more recent times, the advent of the World Wide Web (WWW), has helped create new technologies to promote and provide access to the research performed at NASA. Digital libraries (DLs) are key to this pursuit and provide a means to organize and search for information that would not otherwise be accessible via the web.

The Open Archives Initiative (OAI) has helped to bring together the digital library community and increase the accessibility of scholarly works which exist in digital format. At its heart is the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH), which has become the de facto standard for digital library implementations. The OAI-PMH is a technical specification for a metadata harvesting protocol promoting a standardized interoperability framework among archives. This interoperability addresses such factors as metadata formats, document models, and access protocols. The emergence of the OAI is a sign of the maturation of the science of digital libraries and facilitates the creation and interoperability of digital libraries.<sup>1</sup>

The OAI-PMH defines 2 functional roles; data providers and service providers. Data providers facilitate the exposure of record metadata in a given repository through OAI-PMH formatted queries. Service providers use special metadata harvesters to make these queries and mine data providers. In turn, the service providers process the metadata and provide services to the end user. An OAI digital library search engine browser interface would be a good example of such a service provider.

The OAI-PMH is being widely adopted as the standard for digital library metadata interchange. Approximately six million metadata records were available through the OAI-PMH as of May 2002. The largest effort underway is the National Science Digital Library (NSDL), which expects to “serve millions of users and provide access to tens of millions of digital resources” over the next five years. OAI-PMH conformant data providers may be registered with the OAI and added to their online data provider listings. (Appendix D)

The OAI concluded a 16 month evaluation of OAI-PMH Version 1.1 and released a 2.0 version in June of 2002. Important additions include error reporting, enhanced semantics, a provenance schema to document re-exported metadata, a “friends” schema for the listing of other repositories, and the definition of a single schema which now serves a central validation mechanism for all OAI-PMH requests. Additionally, development tools such as Repository Explorer (which aids in the validation of data provider result sets) have been developed by members of the community and upgraded to handle the 2.0 version of the protocol.<sup>2</sup>

To stay in accordance with the OAI-PMH, NASA’s public technical reports servers needed to reflect the changes in the protocol, so that metadata harvesters would still be able to communicate with NASA’s metadata repositories. NASA’s DL resources consist of several data providers, several of which are complimented by a unique service provider, as well as the NASA Technical Reports Server (NTRS), which serves as the “umbrella” service provider of all the NASA data providers. While evaluating the task of upgrading NASA’s DLs, it was also decided that it would be important to add data provider functionality to NTRS making it an aggregator for the metadata which it harvests from the various NASA repositories.

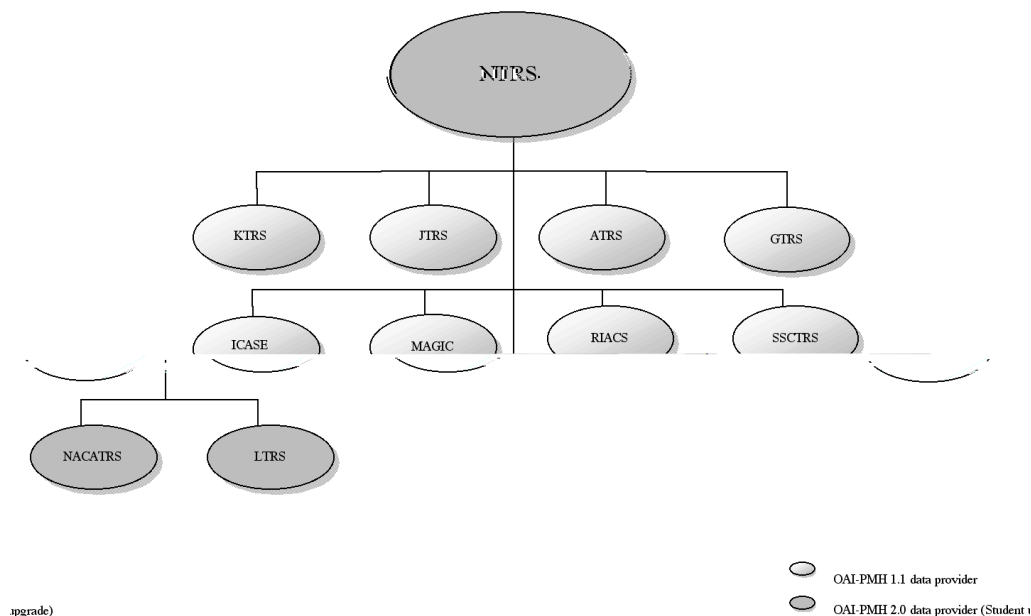


fig 1. The NTRS Umbrella. (note: NTRS is both a data and service provider)

## **Project:**

The aim of this project has been to upgrade several of NASA's OAI data providers so they are compliant to the new OAI-PMH. This took place in three loosely defined stages: preparation, initial development, and the big task.

Though possessing some familiarity with the OAI (1.1) protocol, background readings, as well as a study of the new protocol (released days before the start of my internship) and its implementation guidelines were necessary. Initial readings established the feel of the protocol, while the actual spec and its implementation guidelines served as critical assets during the actual development phase. Setup also included configuring a newly purchased iMac, building a Linux box, and installing all the tools that I would need. Pico was selected as the text editor for the project and installed on several of the bare bones servers that are the homes to NASA's digital library (DL) archives.

With preparations complete, it was time to tackle the task of upgrading each repository. Beginning with the Langley Technical Reports Server (LTRS), a "template of code" was developed to facilitate the new implementations. Once the work on LTRS was tested (using Repository Explorer in the final phase) to the satisfaction of the student and the mentor, the repository was submitted online, for conformance testing with the OAI. After fixing bugs that were discovered, the new LTRS passed the conformance test and was added to the public list of OAI-PMH (2.0) compliant data providers. News of this was announced to the OAI community of data provider implementers. Feedback was received from around the world, resulting in addition changes to accommodate some of the suggestions received. The implementation cycle was then repeated for the NACA Technical Reports Server. At this point, the template was working extremely well, so it was decided to upgrade an additional data provider, the Open Video Project, even though that was beyond the scope of the original project.

After three successful OAI-PMH data provider implementations, it was now time to turn to the big task: transforming the NASA Technical Reports Server (NTRS) from a service provider to a metadata aggregator by adding OAI-PMH data provider functionality. This work required all of the experience gained from the prior upgrades. The task at hand was complex. Harvested metadata resides in NTRS in both OAI-PMH 1.1 and 2.0 formats. Regardless, 2.0 compliant data provider responses would have to be issued. Additionally, the metadata comes from harvests of not one, but several repositories (upstream), and to act as a good aggregator, NTRS would have to be able to identify each unique repository set.

Execution of this task started with adding data provider capabilities to just one of the harvested repositories. After aggressive testing, the others were added as well. With well-formedness issues ironed out it was time once again for validation testing. This was not possible with Repository Explorer (a testing service provided on the Web) because this new version of NTRS was located behind NASA's firewall. To accommodate this, I exported the entire service to another (web accessible) server and was permitted to conduct the validation testing.

**Results:**

The final impact of my work has been the successful upgrade of three OAI data providers and the transition of the NASA Technical Reports Server from a role of service provider to that of aggregator. This project has been delivered to my mentor and the new NTRS will be made public pending NASA's public release procedures. The Langley Technical Reports Server, NACA Technical Reports Server, and the Open Video Project Server are all fully operational and accessible via the web. While focusing on the distribution of NASA's public digital resources, the development of effective metadata harvesting has great implications for widespread application to all digital resources.

## Bibliography

1. “The Open Archives Initiative: Building a low-barrier interoperability framework” . Carl Lagoze and Herbert Van de Sompel, 2001. Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Roanoke VA, June 24-28, 2001, pp. 54-62.  
[draft at <http://www.cs.cornell.edu/lagoze/papers/oai-jcdl.pdf>]
2. “Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative”. Herbert Van de Sompel and Carl Lagoze, 2002. 6th European Conference on Research and Advanced Technology for Digital Libraries, September 2002, Rome.  
[draft at <http://www.openarchives.org/documents/ecdl-oai.pdf>]

## Appendix A (Project Related URLs)

- NTRS: Service provider (User Interface)  
<http://ntrs.nasa.gov/>
- NTRS: Paper describing it  
<http://techreports.larc.nasa.gov/ltrs/PDF/aiaa-95-0964.pdf>
- NTRS: Data Provider  
( Note: Latest NTRS release pending public release )
- LTRS: Paper describing it  
<http://techreports.larc.nasa.gov/ltrs/PDF/tm109162.pdf>
- LTRS: Data Provider OAI-PMH 2.0 interface (base URL for harvesting)  
<http://techreports.larc.nasa.gov/ltrs/oai2.0/>
- NACATRS: Description  
<http://techreports.larc.nasa.gov/ltrs/PDF/1999/tm/NASA-99-tm209127.pdf>
- NACATRS: Data provider OAI-PMH 2.0 interface (base URL for harvesting)  
<http://naca.larc.nasa.gov/oai2.0/>
- Open Video Project: Paper describing it  
<http://ils.unc.edu/~geisg/info/Jnca0112.pdf>
- Open Video Project: Data provider OAI-PMH 2.0 implementation (base URL for harvesting)  
<http://www.open-video.org/oai2.0/>

## **Appendix B (URL formatted calls to data providers)**

Sample Calls (usually made by machine harvester / results format in XML)

### **NASA Technical Reports Server**

Currently NTRS calls cannot be made outside the NASA/Langley firewall, until public release of NTRS 2.0

<http://ntrs.nasa.gov/?verb=Identify>

[http://ntrs.nasa.gov/?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://ntrs.nasa.gov/?verb=ListIdentifiers&metadataPrefix=oai_dc)

[http://ntrs.nasa.gov/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&set=31](http://ntrs.nasa.gov/?verb=ListIdentifiers&metadataPrefix=oai_dc&set=31)

[http://ntrs.nasa.gov/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&from=2002-04-15](http://ntrs.nasa.gov/?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2002-04-15)

<http://ntrs.nasa.gov/?verb=ListSets>

[http://ntrs.nasa.gov/?verb=ListRecords&metadataPrefix=oai\\_dc](http://ntrs.nasa.gov/?verb=ListRecords&metadataPrefix=oai_dc)

[http://ntrs.nasa.gov/?verb=GetRecord&metadataPrefix=oai\\_dc&identifier=oai:naca.larc.nasa.gov:1948:naca-rm-e8i01a](http://ntrs.nasa.gov/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:naca.larc.nasa.gov:1948:naca-rm-e8i01a)

<http://ntrs.nasa.gov/?verb=ListMetadataFormats>

### **Langley Technical Reports Server**

<http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=Identify>

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc)

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&set=31](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc&set=31)

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&from=2002-04-15](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2002-04-15)

<http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListSets>

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListRecords&metadataPrefix=oai\\_dc](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListRecords&metadataPrefix=oai_dc)

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=GetRecord&metadataPrefix=oai\\_dc&identifier=oai:ltrs.larc.nasa.gov:NASA-97-tm110316](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:ltrs.larc.nasa.gov:NASA-97-tm110316)

<http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListMetadataFormats>

### **NACA Technical Reports Server**

<http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=Identify>

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc)

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&set=31](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc&set=31)

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&from=2002-04-15](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2002-04-15)

<http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListSets>

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListRecords&metadataPrefix=oai\\_dc](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListRecords&metadataPrefix=oai_dc)

[http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=GetRecord&metadataPrefix=oai\\_dc&identifier=oai:naca.larc.nasa.gov:1948:naca-rm-e8i01a](http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:naca.larc.nasa.gov:1948:naca-rm-e8i01a)

<http://techreports.larc.nasa.gov/ltrs/oai2.0/?verb=ListMetadataFormats>

### **Open Video Project**

<http://www.open-video.org/oai2.0/?verb=Identify>

[http://www.open-video.org/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://www.open-video.org/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc)

[http://www.open-video.org/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai\\_dc&from=2002-04-15](http://www.open-video.org/oai2.0/?verb=ListIdentifiers&metadataPrefix=oai_dc&from=2002-04-15)

<http://www.open-video.org/oai2.0/?verb=ListSets>

[http://www.open-video.org/oai2.0/?verb=ListRecords&metadataPrefix=oai\\_dc](http://www.open-video.org/oai2.0/?verb=ListRecords&metadataPrefix=oai_dc)

[http://www.open-video.org/oai2.0/?verb=GetRecord&metadataPrefix=oai\\_dc&identifier=oai:www.open-video.org:381](http://www.open-video.org/oai2.0/?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:www.open-video.org:381)

<http://www.open-video.org/oai2.0/?verb=ListMetadataFormats>

## Appendix C (Sample Result Set - formatted in XML)

Sample results from OAI-PMH queries to 2.0 compliant data providers

URL: <http://techreports.larc.nasa.gov/ltrs/oai2.0/?method=Identify>

```
<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-08-07T15:24:32+00:00</responseDate>
  <request>http%3A%2F%2Ftechreports.larc.nasa.gov%2Ftrs%2Foai2.0%2Findex.cgi
  </request>
- <Identify>
  <repositoryName>LTRS</repositoryName>
  <baseURL>http://techreports.larc.nasa.gov/ltrs/oai2.0/</baseURL>
  <protocolVersion>2.0</protocolVersion>
  <adminEmail>m.l.nelson@larc.nasa.gov</adminEmail>
  <adminEmail>t.l.harrison@larc.nasa.gov</adminEmail>
  <earliestDatestamp>1900-01-01T12:00:00Z</earliestDatestamp>
  <deletedRecord>no</deletedRecord>
  <granularity>YYYY-MM-DD</granularity>
- <description>
-   <oai-identifier xmlns="http://www.openarchives.org/OAI/2.0/oai-identifier"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier
       http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
     <scheme>oai</scheme>
     <repositoryIdentifier>ltrs.larc.nasa.gov</repositoryIdentifier>
     <delimiter>:</delimiter>
     <sampleIdentifier>oai:ltrs.larc.nasa.gov:NASA-aiaa-97-1017</sampleIdentifier>
     </oai-identifier>
  </description>
- <description>
-   <friends xmlns="http://www.openarchives.org/OAI/2.0/friends/"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/friends/
       http://www.openarchives.org/OAI/2.0/friends.xsd">
     <baseURL>http://naca.larc.nasa.gov/oai2.0/</baseURL>
     </friends>
  </description>
</Identify>
</OAI-PMH>
```



## Appendix D (Related URLs)

### Documentation

- The Open Archives Initiative Protocol for Metadata Harvesting: (Version 2.0)  
<http://www.openarchives.org/OAI/openarchivesprotocol.htm>
- Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting:  
<http://www.openarchives.org/OAI/2.0/guidelines.htm>
- The Open Archives Initiative Protocol for Metadata Harvesting Changes from 1.1 to 2.0  
<http://www.openarchives.org/OAI/migration/htm>

### Tools

- Repository Explorer:  
[http://www.purl.org/NET/oai\\_explorer](http://www.purl.org/NET/oai_explorer)
- OAI repository conformance tester:  
<http://www.openarchives.org/data/registerasprovider.html>

### Resources

- Open Archive Initiative:  
<http://www.openarchives.org>
- OAI registered data providers:  
<http://www.openarchives.org/Register/BrowseSites.pl>

## Appendix E (NASA's data providers)

NTRS - NASA Technical Reports Server (soon to be publicly released beyond firewall)  
LTRS - Langley Technical Reports Server  
NACATRS - NACA Technical Reports Server (1917 - 1958 technical reports)  
ICASE - Institute for Computing Applications in Science & Engineering  
JTRS - Johnson Technical Reports Server  
MAGIC - Managing Access to Grey Literature Collections  
RIACS - Research Institute for Advanced Computer Science  
ATRS - Ames Technical Reports Server (manually salvaged)  
GTRS - Goddard Technical Reports Server (manually salvaged)  
KTRS - Kennedy Technical Reports Server (manually salvaged)  
SSCTRS - Stennis space Center Technical Reports Server (manually salvaged)  
NACATRS - NACA Technical Reports Server (1917 - 1958 technical reports)  
LTRS - Langley Technical Reports Server

## **Appendix F (LaRC equip**